Which American Languages are Dying? Quantifying the Demographic Vulnerability of Indigenous Languages in the United States

Stephen Cranney

EXTENDED ABSTRACT

INTRODUCTION

Age structure is often invoked in discussions about the vulnerability of endangered languages. Specifically, cohort replacement mechanisms play out as the new dominant tongue is taught to children in school systems and at home while speakers of the declining language age and eventually die (perhaps most prominently in Abrams and Strogatz 2003; see also the UNESCO report on the issue [2003]).

However, despite the common use of demographic terminology and concepts in the endangered language literature, no research has yet employed rigorous demographic methods to precisely quantify vulnerability due to the age structure of the speaking community. To explore how standard demographic methods can play a role in the literature on linguistic vulnerability, here we draw on the concept of population momentum to provide a parsimonious, substantive, and easily interpretable measure: the amount that the respective population would continue to grow or decline if it immediately began exhibiting replacement-level rates, to objectively and comparatively measure the demographic vulnerability of major indigenous languages in the United States.

The two most recent years (2016 and 2017) of American Community Survey data are used to derive the age structure of the seven indigenous language communities in the United States that are best represented in the ACS: Inupiaq, Central Yupik, Western Apache, Navajo, Lakota, Eastern Keres, and Cherokee. Inupiaq and Yupik are indigenous Alaskan languages. Besides Navajo, Yupik has the largest community of indigenous speakers in the Americas today.

Western Apache is spoken in Arizona, and Navajo in both Arizona and New Mexico, with Navajo as the largest indigenous American language community in the United States by far. Lakota is based in the Dakotas while Eastern Keres is spoken predominantly in New Mexico, and Cherokee speakers are based out of the Cherokee tribal populations in North Carolina, Oklahoma, and other areas.

METHODS

Calculation of population momentum requires three derived distributions (Preston, Heuveline, & Guillot 2001, 161-7). The first is *cs(a)*, the proportionate age distribution of the eventual stationary population that will form once the replacement-level vital rates have had enough time to work their way through the age structure of the population. While mortality rates undoubtedly vary from indigenous group to indigenous group, the final stationary population is itself a hypothetical construct that forms after rates have changed; therefore, the most recent publicly available US life table (from the 2014 National Vital Statistics Report) is used to construct *cs(a)*. This choice also has implications for the conceptual interpretation of the population momentum figure: the population momentum represents the proportion increase or decrease of the population if mortality rates were lowered to match those of the baseline US population *and* fertility rates were commensurately changed to replacement level. Therefore, the numbers derived from these rates are meant to construct a standardized hypothetical measure more than serve as a concrete, precise prediction for the future. This standard measure also lends itself to an easily comprehensible conceptual interpretation: how much the language-speaking population would grow if its mortality and fertility rates immediately changed to replacement-level (which would either be an increase or a decrease).

Second, the weight component—*w(a)*—is derived from taking the expected lifetime births above age *a* as the numerator with the mean age at birth in the stationary population as the

denominator. In practice the shape of this function does not vary significantly from location to location even in cases of significant differences in fertility and mortality regimens. Here once again the latest age-specific fertility rates for the US are used (2015), and are comparable to other published rates.

The *w(a)* and *cs(a)* distributions represent baseline components that do not vary from case to case; the *c(a)* distribution is the actual proportionate age distribution of the language population. The language population is measured by a series of questions in the ACS: "does this person speak a language other than English at home?" and "What is this language?" The age structure of the respective linguistic groups are derived from simple weighted frequency tabulations for age by each respective group. As noted, in order to 1) obtain a large enough sample size to accurately gauge the age structure of these indigenous language groups, and 2) use the most recent data, the last two years of American Community Survey data are pooled and used. Also, in 2016 the coding scheme for languages in the ACS changed, so the past two years are relied on for purposes of consistency. The specific languages analyzed here represent all indigenous North American languages that yielded more than 500 unweighted respondents in the pooled two-year ACS sample.

The respondent fills in the blank, and answers are thereafter systematically assigned to one of the established *Ethnologue* linguistic categories. Because the 0-4 age bracket is indefinable for the LAN variable (it is only asked for individuals five and older), the value here is imputed as a factor of the 5-9 age bracket, with the factor itself mirroring the multiplicative relationship between the 5-9 age bracket and the 0-4 age bracket for the ACS sample as a whole. This last imputation introduces the possibility of additional error since the difference in size between these two age brackets likely varies somewhat between groups; however, as this bracket

only represents about 6% of the total population, any error it introduces should be negligible overall. After these three terms are derived they are used to calculate the estimated population momentum using equation 1.

$$Eq.1$$

$$\sum w(a) * c(a)/cs(a)$$

RESULTS

Contrary to conventional wisdom, some indigenous languages are either demographically expanding or are holding steady. Specifically, Central Yupik (considered "vulnerable" by UNESCO) has a population momentum above 1 (1.02—Table 1)— comparable to North America's 1997 momentum of 1.1, and Eastern Keres (considered "definitely endangered") has a population momentum of .98 (Table 1) based on its current age structure. For the Yupik, their score of 1.02 means that they would grow to 102% of their current population before their age structure settled into its long-run equilibrium. However, the rest of the indigenous groups analyzed here have contracting population structures. The Inupiaq ("severely endangered"), the Lakota ("critically endangered"), and the Navajo (by far the largest indigenous language group and perhaps the one with the most resources devoted to language preservation—although still characterized as vulnerable) all have population momentums in the .8 range (Table 1). As a point of comparison, the population in Germany in 1997 was .88 (Preston & Guillot 1997), and the population of Japan in 2000 was around .9 (Blue & Espenshade 2011). The Cherokee ("definitely endangered") and the Western Apache ("definitely endangered") have even more severe negative population momentums at .74 and .58, respectively (Table 1). Once again, this means that if the Western Apache mortality rates shifted to the general US level at the same time their fertility rates shifted to replacement level, they would still shrink to nearly half of their

current population before their population stabilized. Blue and Espenshade's 2011 survey of the population momentum histories of 27 countries, none of them reached anywhere in the .7 levels or below.

Table 1

Languages	Population Momentum	SE	N[1]
Inupiaq	0.83	0.006	850
Central Yupik	1.02	0.0005	2900
Western Apache	0.58	0.001	600
Navajo	0.82	0.0008	11500
Lakota	0.86	0.005	750
Eastern Keres	0.98	0.009	850
Cherokee	0.74	0.008	650

Source: U.S. Census Bureau, American Community Survey, 2016 and 2017.
For information on sampling and nonsampling error, see
< https://www2.census.gov/programs-surveys/acs/tech_docs/accuracy/ACS_Accuracy_of_Data_2016.pdf >
1. Rounded for disclosure avoidance.

Here we provide one example of how demographic methods can be fruitfully used in the linguistic endangerment literature to supplement already existing quantitative indicators of linguistic endangerment (such as the Expanded Generational Disruption Scale (EGDS) and the UNESCO Degree of Endangerment) that largely rely on qualitative judgment calls. We also explore the possibility of incorporating other standard demographic measures such as net reproductive rate and age-specific rates of growth or decline into our overall picture of demographic vulnerability.

REFERENCES

Abrams, D. M., & Strogatz, S. H. 2003. Linguistics: modelling the dynamics of language death. *Nature*, *424*(6951), 900.

Arias, E., Heron, M., and Xu, J. 2014. United States life tables, 2013. *National Vital Statistics Reports 66*(3).

Martin, J., Hamilton, B., Osterman, M., Driscoll, A., and Mathews, T. 2017. Births: final data for 2015. *National Vital Statistics Reports 66*(1).

Moseley, C. (ed.). 2010. *Atlas of the World's Languages in Danger*, 3rd edn. Paris, UNESCO Publishing. Online version: http://www.unesco.org/culture/en/endangeredlanguages/atlas

Preston, S., and Guillot, M. 1997. "Population dynamics in an age of declining fertility." *Genus 53*(3-4), 15-31.

Preston, S., Heuveline, P., and Guillot, M. 2001. *Demography: Measuring and Modeling Population Processes*. Oxford, UK: Blackwell Publishers.

Simons, G.F. and Fennig, C.D. (eds.). 2018. *Ethnologue: Languages of the World, Twenty-first edition.* Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.